AD-A258 405

Semiannual Progress Report to ONR
May 1, 1992 - October 31, 1992

Office of Naval Research
Bioacoustic Signal Classification
Cognitive and Neural Sciences
& Mathematics Division

Classification of Complex Sounds
#ONR N00014-92-J-1164

Bruce G. Berg
Principle Investigator
Department of Cognitive Sciences
School of Social Sciences
University of California, Irvine
Irvine CA 92717

92-30558

**WORK IN PROGRESS**

## Computational models of spectral discrimination: Spectral weights and internal noise.

We have identified at least two factors which characterize listeners' abilities to discriminate complex spectra. One is the ability to integrate information across different auditory channels. Using COSS analysis we have found that good listeners show spectral weights that are close to optimal, whereas poor listeners show weights which deviate from optimal weights. A second factor is internal noise, that is, the inherent variability of listeners. We have developed a model which incorporates both two factors. An important aspect of this model is that it has no free parameters. Estimates of spectral weights and internal noise are completely determined by a trial-by-trial analysis of the data. Predictions are very consistent with thresholds obtained in a profile analysis task. Given this initial success, we are confident that the model will become a valuable tool. In addition to identifying the characteristics of superior listeners, the model is being used to investigate training effects. For instance, are the improvements in performance following training attributable to a reduction in internal noise or to a more optimal strategy for integrating spectral information (e.g. better spectral weights)?

[A report of this work will be given at the Meeting of the Acoustical Society of America, Oct 31, New Orleans].

## Spectral weights and the reliability of spectral information

Spectral weight estimates obtained with COSS analysis provide an assessment of how listeners attend to and use spectral information in order to discriminate complex sounds. Detailed knowledge about how an individual uses spectral information may be useful in devising training strategies aimed at increasing the efficient use of available information. One approach to this problem is to understand what conditions lead to or induce a change in spectral weights. In a preliminary study (one subject), we have found that spectral weights are dependent on the variability of spectral information. If the variability of a given spectral component is increased (decreasing the reliability of information), the spectral weight associated with that component decreases, whereas if the variability of the same component is decreased (increasing the reliability of information), then the weight for that component increases. In other words, these data show that the listener is able to monitor the reliability of information provided by different components and then attend to those channels providing the most reliable information. We will soon begin a systematic study of this issue.

## Spectral discrimination and envelope cues

As discussed in other sections of this report, we have found evidence that listeners use cues related to the envelope of the temporal waveform to discriminate complex, narrow band spectra. We have recently completed data collection in an experiment designed to determine the "effective bandwidth" for envelope cues. The task is analogous to the classic study by Fletcher (1940). If we plot the energy required to detect a tone in white noise as a function noise bandwidth, the function increases up to a certain point and then remains constant. The "breakpoint" of this function is a rough estimate of critical bandwidth, that is, the width of the auditory filter centered at the frequency of the tone. Similar functions are obtained in our experiment. We begin with a three tone complex with frequencies of 980 Hz, 1000 Hz, and 1020 Hz. For the standard, all three tones have the same intensity, and for the signal, the level of the 1000 Hz component is increased. Threshold is defined as the level of this increment which is necessary for 71% correct discriminations between the two sounds. In the second condition, the standard consists of two additional tones (with the same intensity) having frequencies of 960 Hz and 1040 Hz, and in additional conditions, we continue adding tones at both ends of the spectrum, each spaced 20 Hz apart. The signal component is always 1000 Hz. As in the Fletcher task, we are interested in thresholds as a function of stimulus bandwidth. Our task differs from Fletcher's in two important ways. First, we use a roving level procedure in which the level of each stimulus presentation is a random variable with a 20 dB range -- this degrades information related to absolute intensity. Second, the digital-to-analog conversion rate is changed (in a random fashion) prior to each stimulus presentation -- this degrades information related to pitch. In theory, the only cue available for discriminating the two sounds is the envelope of the temporal waveform, which we believe is related to timbre.

The results are very similar to "Fletcher-type" functions. Generally, thresholds first increase as the number of components increase. Beyond a certain bandwidth (number of components), however, thresholds either remain constant or actually decrease. We call this breakpoint the "critical bandwidth for envelope cues". In most cases this bandwidth is three to four times wider than traditional estimates of the critical band. In other words, processing temporal information appears to involve auditory filters with much wider bandwidths than those used to process the spectral energy of a sound. Of additional interest is the large variation among individual listeners -- some listeners show threshold functions which "break" at relatively narrow bandwidths, whereas the threshold functions for other listeners continue to increase across bandwidths as large as several thousand hertz. The former type of listeners appear to be able to extract envelope information from a relatively narrow band of frequencies, even if those frequencies are embedded within a much wider spectrum. Furthermore, they often have thresholds that are 10 dB lower than listeners which show no breakpoint in their threshold functions.

There are two important implications of this work. First, recent evidence from a number of labs strongly suggest that intensity cues are relatively unimportant for discriminating complex spectra and an energy detection model probably cannot account for listeners' abilities to discriminate complex sounds. Currently, we are using the data from our study to develop alternative models to guide our research, specifically, one which quantifies cues based on timbre. The second aspect of this work concerns the large individual differences that we have found in this task. Apparently, different listeners process the available information in different ways. The second stage of this research involves using a COSS analysis with the aim of gaining more detailed knowledge about these individual differences.

## COMPLETED RESEARCH

Perhaps my most important work during the last three years is the development of COSS analysis, an empirical technique which provides a detailed, "microscopic picture" of how listeners process spectral information when discriminating complex sounds. The technique has great potential, as evidenced by the important advances and insights that we have thus far gained using COSS analysis.

In essence, the technique consists of adding a small, independent intensity perturbation to the level of each spectral component of a complex stimulus during each trial of a discrimination task. Using COSS analysis, we determine the relative importance or weight of each component by examining the effects that these level perturbations have on a listener's responses. If a component is relatively important, making a large contribution to the listener's decisions, than level perturbations added to that component have a relatively large effect on responses. On the other hand, if a component is relatively unimportant, then any level perturbations added to that component have relatively little effect on responses. COSS analysis is used to quantitatively assess the effects of level perturbations, yielding a set of spectral weights that reflect the relative importance of each spectral component of a complex stimulus. Technical details about the mechanics of the analysis, per se, are found in Berg (1889) and Berg and Green (1990). Here, we focus on the end results of this analysis and several theoretical interpretations of weight estimates.

The investigative power of the technique is nicely demonstrated by the following experiment. The amplitude spectra for two complex stimuli are shown at the top of Fig. 1. Each consists of eight tones. The tones are equally spaced with respect to log frequency and range from 200 Hz to 5000 Hz. In theory, each of the tones occupy a different auditory channel or critical band. The **flat spectrum** shown on the right is the standard and the **rippled spectrum** shown on the left is the signal. The listener's

task is to discriminate these two spectra. We define the threshold as the "depth of the ripple" necessary for 71% correct discriminations.

We are primarily interested in how listeners integrate auditory information across channels. In order to degrade cues based on absolute intensity, the overall level of each stimulus presentation is sampled from a uniform distribution with a mean of 60 dB SPL and a range of 20 dB (i.e. random rove procedure). The reason for using this procedure is that, in theory, listeners make spectral shape discriminations based on an across channel integration of relative intensity when absolute intensity cues are degraded (Green, 1988). [Note that the COSS analysis technique provided the first direct evidence supporting this theory (Berg and Green, 1990).] A normative model of this integrative process is simply a weighted sum of the auditory channel outputs, $\Sigma a_i x_i$, where $x_i$ is the level output from the $i^{th}$ channel and $a_i$ is the weight associated with that channel. For the task illustrated in Fig. 1, the **optimal weights** are +1 for the tones (or channels) that are incremented in level relative to the standard and -1 for the tones which are decremented in level relative to the standard. The weighted sum is used as a decision statistic; the greater the value of this statistic, the more evidence we have that a signal was presented.

COSS analysis is used to estimate the channel or spectral weights used by listeners. Results for three listeners are shown in the panels at the bottom of Fig. 1. In each panel, optimal weights are shown by solid symbols and estimated weights are shown by open symbols connected by lines. For convenience, a dotted horizontal line is added at zero weight. All three listeners are similar in that they tend to use information from only a limited portion of the spectrum; two or three weights are relatively large, either negative or positive in sign, and the remaining weights are close to zero. Our interpretation of zero weight is that the corresponding tone has little influence on a listener's discriminations. Although all subjects appear to use a limited portion of the available spectral information, an important point illustrated by these data is the striking differences across individuals. One listener uses information from the three central tones of the spectrum (right panel), another rakes level comparisons of the two lowest frequency tones (middle panel), and the third appears to compare the level of a central tone with high and low frequency tones at the spectral edges (left panel).

There is no current theory offering an explanation for individual differences. Differences could be attributable to sensory processes, attentional processes, or differences in training, history and experience. To fully appreciate the usefulness and potential of COSS analysis for investigating these issues, however, it is informative to consider the level of analysis that we would possess without the benefit of the COSS technique. The numbers listed in the panels represent thresholds (the decibel level of the increment or decrement of each tone of the signal

relative to the standard). Note that the three listeners have essentially identical thresholds. On the basis of these data, we would conclude that there are **no** differences across listeners!

A consistent finding that we have observed across many studies is that a listener's discrimination performance is highly related to estimates of spectral weights -- "good" listeners show weights that are close to optimal, whereas "poor" listeners show weights that deviate greatly from optimal. The chief point here is that we now possess the means for a highly detailed assessment of individual listeners' abilities and tendencies for using spectral information in various listening tasks [A modification of the technique also allows an assessment of how listener's use temporal information (Berg, 1990) or temporal-spectral information (Dai and Berg, 1992)]. This procedure is well suited for investigating a number of issues related to individual differences: 1) Using COSS analysis to distinguish superior and inferior listeners, we might determine the degree of transfer to different auditory tasks. That is, are superior listeners more efficient at integrating auditory information across tasks? 2) Using spectral weights as a training aid -- an assessment of an individual's tendencies for using acoustic information might be a useful tool for devising training strategies with the aim of increasing the efficient use of the available information. 3) If spectral weights reflect an individual's "natural listening tendencies", it might be possible to exploit those tendencies by "tailoring" the sound. In other words, spectral weights may be useful in developing signal enhancement techniques based on psychological aspects of the listener (providing a complement to traditional signal enhancement techniques based on physical aspects of sound with the listener "removed from the loop"). Currently, we are exploring some of these issues by investigating conditions which effect changes in a listener's weights.

Another important development is the use of COSS analysis for devising and testing theories about auditory processes and computational mechanisms. Recent findings from a number of researchers make it increasingly apparent that traditional models, based on energy cues, sometimes cannot characterize listener's abilities to detect subtle spectral changes among sounds. One would not be too far afield to propose that new models must necessarily be devised to further guide our research. We are beginning to recognize COSS analysis as an invaluable tool for theory development and empirical testing of such models.

Work done in collaboration with David Green shows that pitch and timbre cues are important for discriminating spectral changes in complex sounds. Briefly, consider a profile task in which listeners discriminate a standard consisting of three tones of equal intensity (i.e. 200 Hz, 1000 Hz, and 5000 Hz) from a signal for which the intensity of the central tone is incremented (producing a "spectral bump"). As before, a roving level procedure is used to degrade absolute intensity cues. Traditional energy-based models (see Durlach, et al., 1986; Berg

and Green, 1990) assume that listeners make simultaneous level comparisons across channels. If the level of the 1000 Hz tone is greater than the mean level of the two nonsignal tones, then we have evidence that the sound has a relative intensity increment at 1000 Hz. This computational model is supported by estimates of spectral weights. The weights for the two nonsignal components are close to -0.5, relative to unit weight for the signal component. Now consider the predictions of this model when the sounds have a very narrow bandwidth, for example, when the two nonsignal components have frequencies of 990 Hz and 1010 Hz, respectively. In this case, since all three tones occupy the same critical band, a comparison of across channel level differences is no longer viable. Supposedly, this limits energy based cues to the output of a single channel. Given the 20 dB range in absolute intensity, we would expect thresholds to be extremely high. Furthermore, since tones are "unresolved" within a critical band, we also expect that spectral weights for all three components will have the same sign and roughly the same magnitude. Our data show that neither of these two predictions are correct. Both thresholds and spectral weight estimates are essentially the same for broad band and very narrow band conditions.

Our discovery that spectral weights for unresolved tones have different signs and magnitudes was intriguing and initially unsettling. How can unresolved tones have different weights? Following several additional studies, we proposed two models to account for both thresholds and spectral weight estimates. One model considers pitch cues and the other considers timbre.

The potential use of pitch cues was investigated by making the digital-to-analog conversion rate (DAC) a random variable. This procedure degrades cues based on pitch much in the same way as a roving level procedure degrades absolute intensity cues (Richards, et al. 1989). We found that when the stimulus bandwidth is 80 Hz (960 Hz, 1000 Hz, and 1040 Hz tones), randomization of the DAC rate leads to a 12 dB increase in thresholds, providing evidence that listeners are basing their discriminations on pitch cues. When the bandwidth is 20 Hz (980 Hz, 1000 Hz, and 1010 Hz tones), randomization of the DAC rate has no effect, showing that discriminations are not based on pitch cues in this condition.

Spectral weight estimates were also obtained for the two conditions. Results are shown for six observers in Fig. 2 for the 80 Hz bandwidth condition and in Fig. 3 for the 20 Hz bandwidth condition. The profoundly different pattern of weights across the two conditions certainly reflect differences in the computational processes used in discriminating the stimuli. Here we observe another facet of COSS analysis, namely, the ability to acquire detailed behavioral data that uniquely reveal aspects of auditory processes that cannot be observed with traditional means (i.e. threshold data). The power and value of such data is illustrated below with brief descriptions of two models that

predict the pattern of weights for the two conditions.

In conditions for which pitch is used as a discrimination cue, a version of Feth's (1974) **envelope weighted average of the instantaneous frequency** (EWAIF) model provides a reasonable account of the data. The EWAIF model is a theoretical calculation of pitch. Feth and Stover (1987) proposed that calculated EWAIF values can be used to discriminate changes in spectral shape. Two modifications of that model are made. First, rather than using the actual waveform, EWAIF values are calculated from the output of a single auditory filter. Second, rather than using a filter centered on the stimulus, we use a filter with a center frequency that is higher than the highest frequency tone of the complex (or lower than the lowest frequency tone). In other words, the complex sound is filtered through the skirt of a arbitrary filter (additional details can be found in Berg, et al. 1992).

The six panels at the bottom of Figure 2 show the spectral weights from six listeners in the three-tone, 80 Hz bandwidth condition. An unusual, though revealing, aspect of these data is that the weight of greatest magnitude is **not** found for the 1000 Hz signal component, but is instead found for the 960 Hz, nonsignal component. Also note that the other nonsignal component (1040 Hz) is positive in most cases. The panel at the top of the figure show the weights obtained using simulations of the modified EWAIF model. To a first approximation, the model provides a reasonable account of these data. Moreover, the unique pattern of weights observed for this condition allows us we to rule out a number of alternative models.

Evidence suggests that discriminations in the 20 Hz bandwidth condition are made on the basis of timbre cues, that is, increasing the intensity of the 1000 Hz tone makes the signal sound "less rough" than the standard. We believe that this roughness cue is derived from the envelope of the temporal waveform, and we have developed a new model which quantifies this cue (Berg and Green, 1991; Green, et al., 1992). Specifically, the model computes a decision statistic based on the power spectrum of the envelope. Spectral weights obtained from simulations of this model are shown in the upper panel of Fig. 1. Again, to a first approximation, we are able to provide a reasonable account of the data. Recently, we have generalized this model in order to provide an explanation for some puzzling results in a tone-in-noise detection task. Gilkey (1987) and Kidd, et al. (1989) found that using a roving-level procedure in a such a task increases thresholds by only a few decibels, far less than predicted by traditional models based on energy cues. One characteristic of our model is that it is largely unaffected by large variations in absolute intensity (Technical details of the model are found in Green et al., 1992). This property of level independence alone makes the model extremely appealing, since it better characterizes discriminations made in natural listening environments.

Although the potential and practical benefits related to the two models require a longer time-frame than the work concerned with several immediate issues related to individual differences, the development of computational models is nonetheless critical in increasing our understanding of individual differences.  For one, such models are important for understanding and interpreting the behavioral data obtained with COSS analysis.  By analogy, one can view the current state of COSS analysis as a "first look through a crude microscope" -- although it provides an extremely detailed "picture", we are not always certain what exactly it is that we are observing.  We are confident that continued work will assure the refinement of this powerful technique.

## REFERENCES

Berg, B.G. (1989).  Analysis of weights in multiple observation tasks.  J. Acoust. Soc. Am., 86, 1743-1746.

Berg, B.G. (1990). Observer efficiency and weights in a multiple observation task.  J. Acoust. Soc. Am., 88, 149-158.

Berg, B.G., and Green, D.M. (1990).  Spectral weights in profile listening.  J. Acoust Soc Am., 88, 758-766.

Berg, B.G. and Green, D.M. (1991).  Spectral shape discrimination of narrow-band spectra.  J. Acoust. Soc. Am., 89, 1911 (abstract).

Berg, B.G., Nguyen, Q., and Green, D.M. (1992).  Discrimination of narrow band spectra: I. Spectral weights and pitch cues.  J. Acoust. Soc. Am., in press.

Dai, H., and Berg, B.G. (1992). Spectral-temporal weights in a profile task. J. Acoust. Soc. Am., 92, 1346-1355.

Durlach, N.I., and Braida, L.D., and Ito, Y. (1986).  Towards a model for discrimination of broadband signals.  J. Acoust. Soc. Am., 80, 63-72.

Feth, L.L. (1974).  Frequency discrimination of complex periodic tones.  Percept. Psychophys., 15, 375-378.

Feth, L.L., and Stover, L.J. (1987).  Demodulation processes in auditory perception.  In W.A. Yost and C.S. Watson (Eds.), Auditory Processing of Complex Sounds (Erlbaum: Hillsdale, NJ), pp 76-86.

Gilkey, R.H. (1987).  Spectral and temporal comparisons in auditory masking. In W.A. Yost & C.S. Watson (Eds.), Auditory Processing of Complex Sounds.  Hillsdale, NJ: Lawrence Erlbaum.  P26-36.

Green, D.M. (1988).  Profile Analysis: Auditory Intensity Discrimination.  New York: Oxford Univ. Press

Green, D.M., Berg, B.G., Dai, H., Eddins, D.A., Onsan, Z., Nguyen, Q. (1992).  Spectral shape discrimination of narrowband sounds. In Review.

Kidd, G., Jr., Mason, C.R., Brantley, M.A., and Owen, G.A. (1989). Roving-level tone-in-noise detection.  J. Acoust. Soc. Am., 86., 1310-1317.

Richards, V.M., Onsan, Z.A., and Green, D.M. (1989).  Auditory profile analysis: Potential pitch cues.  Hear. Res., 39, 27-36.

## Published Papers (Refereed Journals)

Green, D.M., and Berg, B.G. (1991). Spectral weights and the profile bowl. Quarterly Journal of Experimental Psychology, 43A, 449-458.

Dai, H., and Berg, B.G. (1992). Spectral-temporal weights in a profile task. J. Acoust. Soc. Am., 92, 1346-1355.

## Papers in Press (Refereed Journals)

Berg, B.G., Nguyen, Q., and Green, D.M. (1992). Discrimination of narrow band spectra: I. Spectral weights and pitch cues. J. Acoust, Soc. Am., in press.

## Submitted Papers (Refereed Journals)

Green, D.M., Berg, B.G., Dai, H., Eddins, D., Onsan, Z., and Nguyen, Q. (1992). Detecting spectral shape changes in very narrow frequency bands. In review, J. Acoust. Soc. Am.

## Book Chapters

Berg, B.G., and Green, D.M. (1992). Discrimination of complex spectra: Spectral weights and performance efficiency. In Y. Cazals, L. Demany, and K. Horner (Eds.), *Auditory Physiology and Perception. Oxford: Pergamon Press.*

## Invited Addresses and Colloquia

University of California, Berkeley, November, 1991
University of California, Riverside, April, 1992

## Contributed Presentations

Berg, B.G. (1992). Discrimination of narrow band spectra based on pitch cues. Paper presented at the Fifteenth Midwinter Research Meeting of the Association for Research in Otolaryngology, St. Petersburg, FL, February 2-6, 1992. (published abstract p93).

Berg, B.G. (1992). Estimation of multichannel weights in an Auditory Task. Paper presented at the Seventeenth Annual Interdisciplinary Conference, Jackson, Wyoming, January 15, 1992.

Berg (1992). COSS analysis and Individual Differences. Paper presented at the ONR Bioacoustic Signal Classification Workshop, April 15-17, Wrightsville Beach, North Carolina.
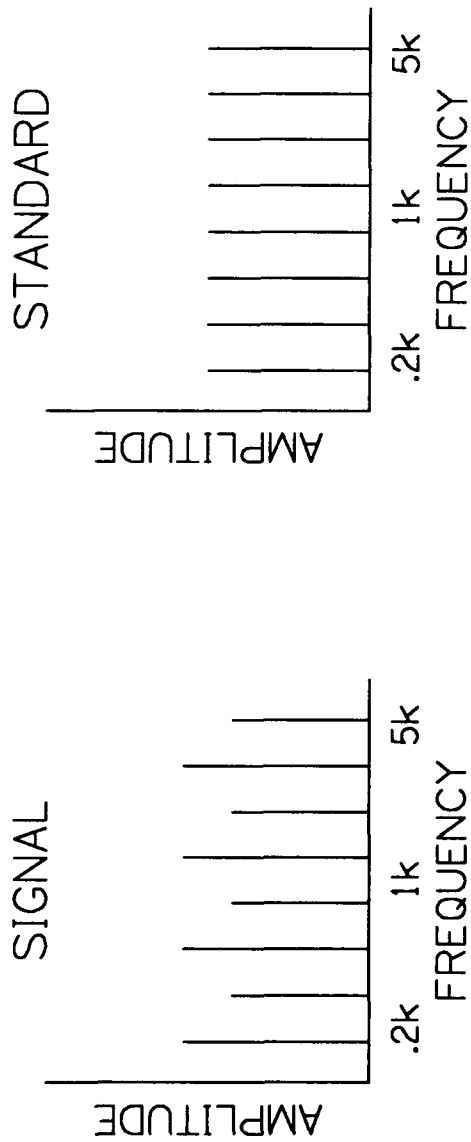
## Book Reviews

Harp, S.A., and Berg, B. (1992). Detection theory for everyman? (Book review - Detection Theory: A User's Guide, by N.A. Macmillan and C.D. Creelman. Cambridge/NY: Cambridge University Press, 1991.) J. Math. Psych., in press.

## Training

Currently, there are two graduate students working in my laboratory. Both were supported by the ONR grant as 100% FTE during the summer of 1992, and both will continue to work and be supported during the 1992-93 school year. Curtis Southworth, a third year graduate student, plans to soon develop a dissertation topic related to COSS analysis. Matthew Turner is a first year student. He accepted the offer from the UC Irvine graduate program with the expressed intent of working in my laboratory. Both students are quite advanced in terms of their mathematical and computer skills, and I am very optimistic about their potential contributions.
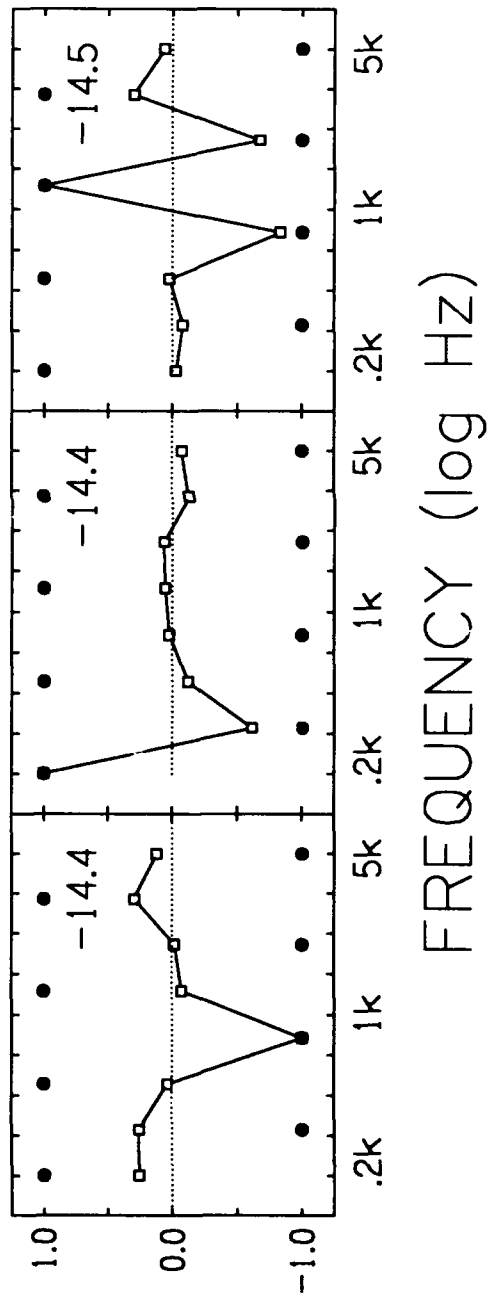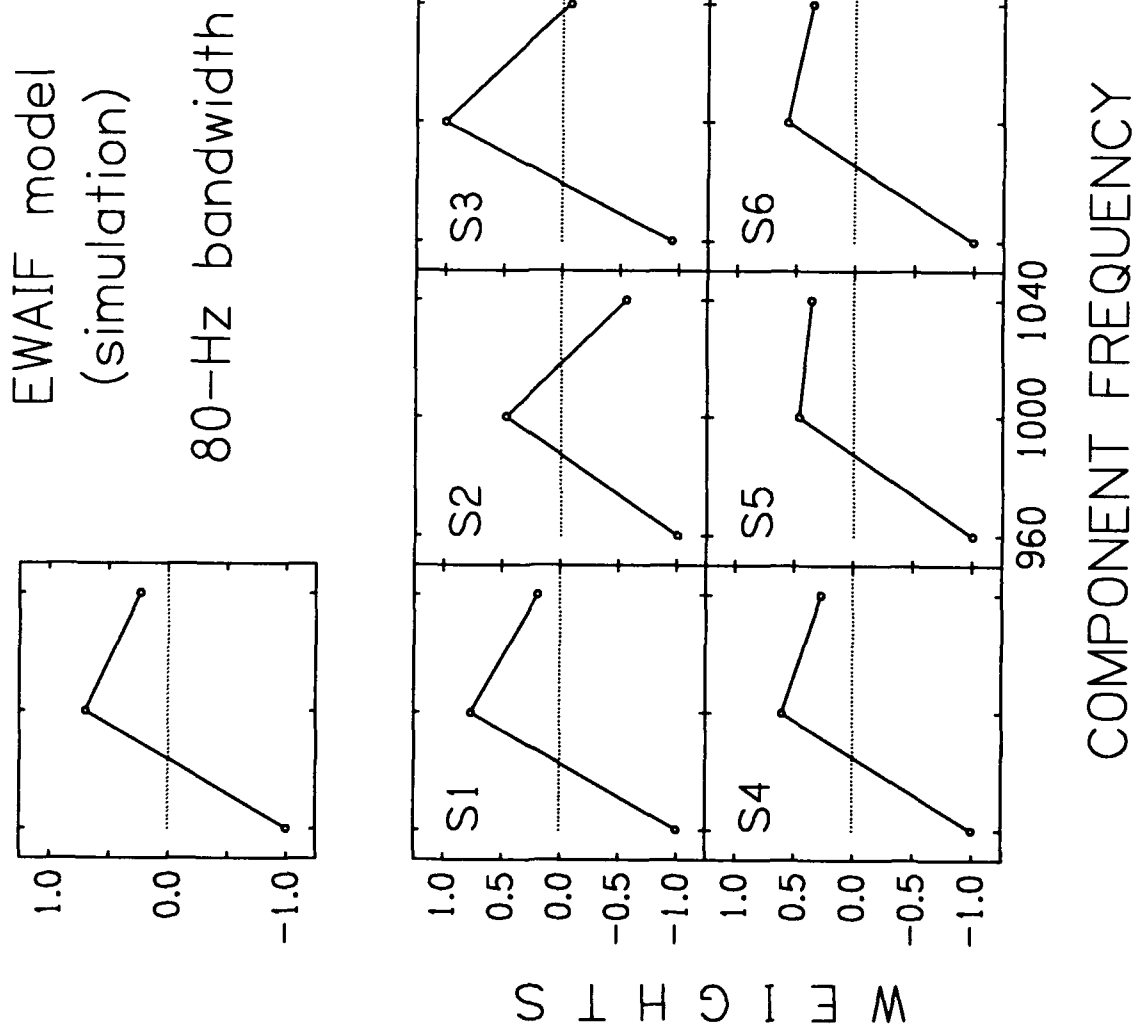
SIGNAL

STANDARD

Spectral Weight Estimates

FREQUENCY (log Hz)

FIGURE 1

EWAIF model
(simulation)

80-Hz bandwidth

COMPONENT FREQUENCY

WEIGHTS

FIGURE 7

# Envelope Spectrum
## (simulation)

### 20-Hz bandwidth



FIGURE 3